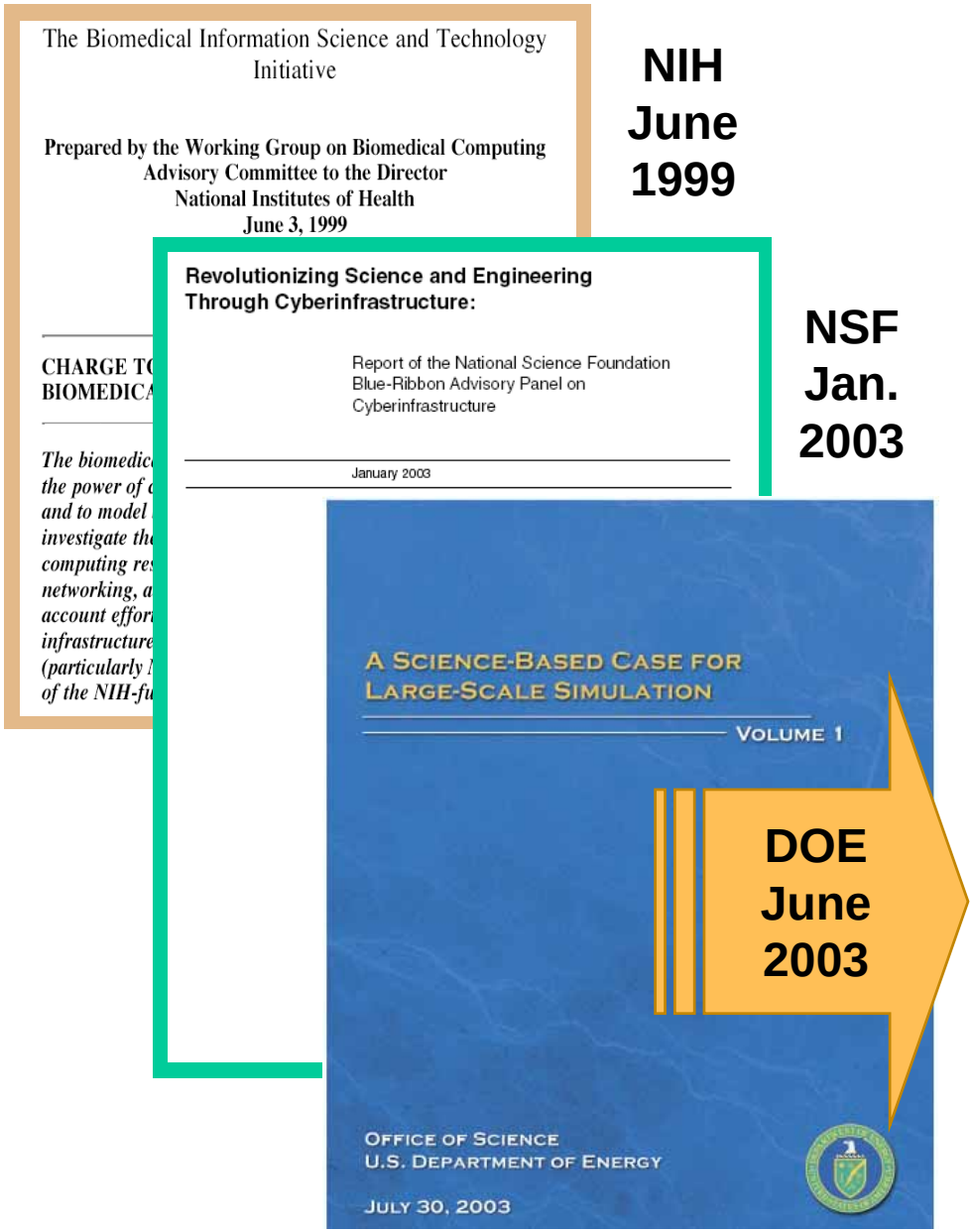
Overview of the BG Consortium

Rick Stevens
Argonne National Laboratory
The University of Chicago

Blue Gene Consortium Kickoff

- Welcome
- Agenda Review
- Overview of the Consortium Goals
- Goals of this meeting

Major Petascale Computing Scientific Goals



- Predicting Future Climates
- Understanding the Behavior of Complex Microbial Systems
- Tuning Flame Chemistry for Efficient Combustion
- Designing Material Properties from the Nanoscale Up
- Bringing a Star to Earth
- Realistic Simulation of Particle Accelerators
- Fundamental Nature of Matter
- Predictive Basis for Cleanup Decisions
- Designing and Optimizing Fusion Reactors
- Understanding the Quark-Gluon Plasma
- …

Our Vision

- Petascale Computing
  - Increase by several orders of magnitude through new architectures the computing power that can be applied to individual scientific problems, thus enabling progress in understanding complex physical and biological systems.

- Grid Computing
  - Interconnect the world's most important scientific databases, computing systems, instruments and facilities to improve scientific productivity and remove barriers to collaboration.

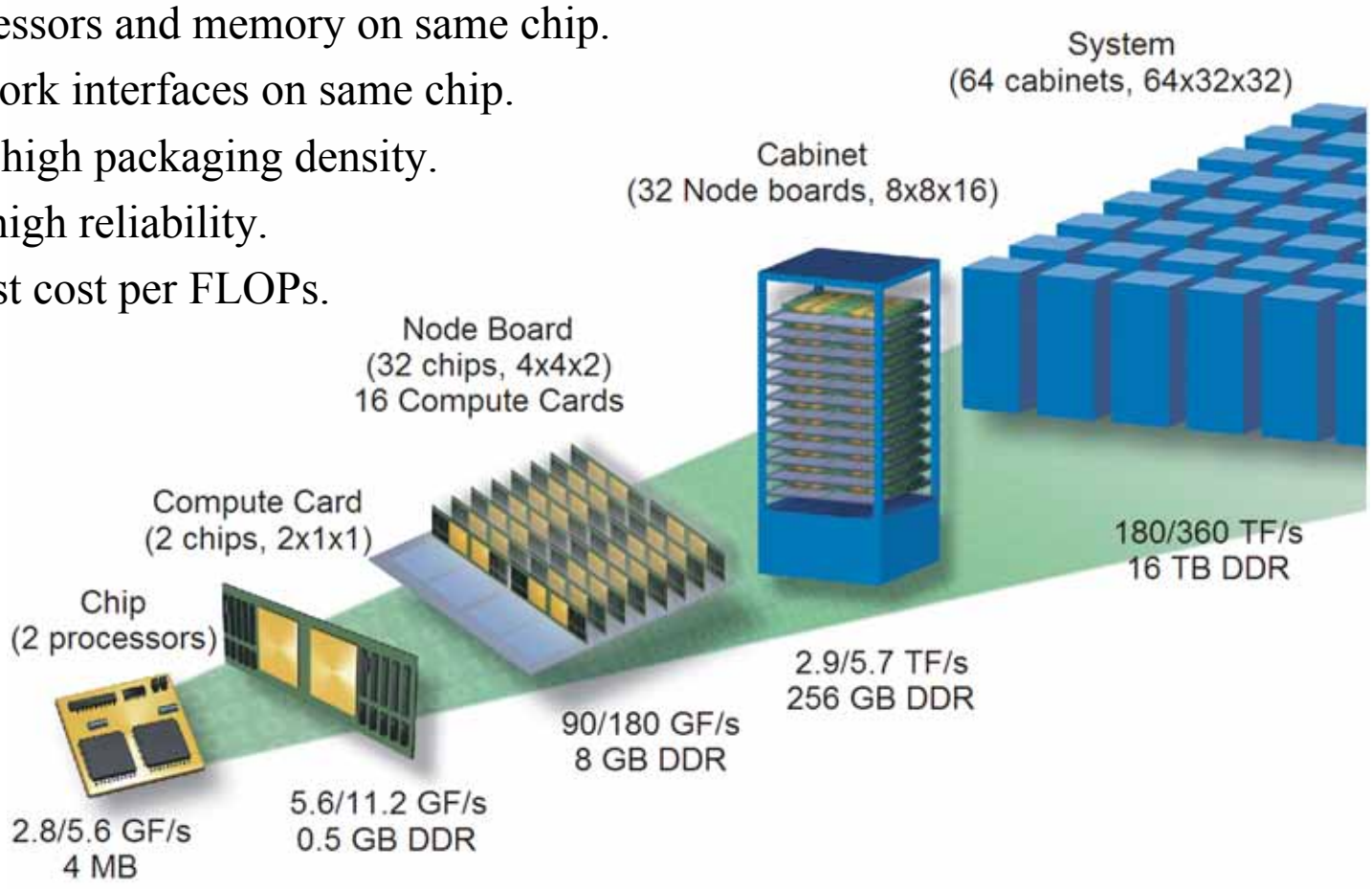
mcs

Why Computational Science Matters

- Safe and abundant food supplies
- Sustainable and benign energy sources
- Effective management of disease and aging
- Novel materials and renewable industrial feedstocks
- Advanced computational devices beyond silicon
- Wide variety of molecular scale machinery
- Self-assembly and self-reproduction technologies
- Understanding our origins and place in the universe

The Blue Gene Family of Computers

- Puts processors and memory on same chip.
- Puts network interfaces on same chip.
- Achieves high packaging density.
- Delivers high reliability.
- Has lowest cost per FLOPs.



Agenda Review

- Welcome
- ANL (overview)
- IBM (BG status review and consortium proposals)
- LLNL (BG project update)
- Q+A
- Breakout Charges
- Lunch
- Breakout session 1 (management, sw, hw)
- Breakout session 2 (management, apps, ff)
- Reporting out
- Agency Roundtable/Panel
- Consortium Kickoff Ceremony
- Next steps and conclusions

mcs

How We Got Here

- ANL and IBM discussions
  - 2002/2003 and finalizing at SC'03 with a plan
  - Document outlining consortium responsibilities
    - Focus on the supporting smaller hardware sites
      - Hub and spoke model to reduce support costs
      - Each participant needs some explicit commitment of effort
  - Needed to find a way to make small systems feasible
    - Lower the cost of ownership through community processes
      - Self help, "open source" like community support
      - Shared effort for training and support
  - Way to provide community feedback to IBM on BG
  - Develop a national BG user community

mcs

The Blue Gene Consortium

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

- Goals
  - Provide new capabilities to selected applications partnerships
  - Provide functional requirements for Petaflops/sec version of BG
  - Build a community around a new class of HPC architecture
    - 30 university and lab partners
    - ~10 HW partners + ~20 SW collaborators
  - Develop a new (sustainable) model of partnership
    - "research product" by passing normal "productization" process/costs
    - Community based support model (hub and spoke)
  - (re-)Engage computer science researchers with HPC architecture
    - Broad community access to hardware systems
    - Enable scalable OS research and novel software research
  - DOE, NSF, NIH, NNSA, IBM partnership
    - CS research, computational science, architecture development
  - Kickoff meeting is 27th April, 2004 in Chicago

mcs

BG/ and Possible Paths to Petaflops

- Potential successor machines to BG/L maybe capable of reaching petaflops/sec performance on some applications.
- One possible goal of the BG Consortium could be to help foster interest in a followon project to BG/L to build a petascale class system (BG/P)
- One goal of the consortium could be to provide sustained (apps and ss) input during the design and development process to improve BG/P
- Another goal might be to develop an applications community able to exploit BG/X class architectures

A Proposed Consortium Model

- Several Ways to Join
  - Commit to acquiring minimal hardware
    - 1 rack is the working assumption today (IBM will talk about price)
  - Commit to porting systems software or tools
    - Libraries, file systems, schedulers, language, etc.
  - Commit to porting application codes
    - Bio, astro, materials, nano, climate, fusion, geo, etc.
  - Contribute technology to future systems
    - Simulators, compilers, design IP etc.
- We think of hardware sites as “partners”
- Non hardware sites as “collaborators”
- Collaborators need to have a Partner “sponsor”
- Both are members of the consortium
- Argonne is willing to sponsor collaborators

mcs

Overview of Consortium Goals

- Build a robust community able to thoroughly evaluate the BG/L architecture
- Porting and developing applications for BG/L
- Assemble a critical mass to port and develop critical systems software to the BG/L architecture
- Provide feedback to IBM on future systems related to BG/L
- Provide an organizational structure to help support the academic and laboratory providers and users of the BG/L system and potential followons
- Support the engagement of CS departments on BG related software projects

mcs

BG Consortium a Third path?

- Can we demonstrate a HPC business model that:
  - Provides research community with access to vendor designed and purpose-built scientific supercomputers
  - Avoids costs and delays of full productization
  - Engages the scientific community directly in the dialogue for future architectures
  - Enables open source tools and software infrastructure
  - Encourages a community self-help support model
  - Reinvigorates the academic community in HPC
  - Builds community and involves next generation researchers

Community Evaluation of BG/L

- Diverse set of users to understand and attempt to use BG/L for important applications
  - Aim is to get 30-50 applications up on BG/L
  - Develop performance and scaling models for each
- Evaluation of:
  - Hardware (CPU/network structure)
  - Programming model (with limitations)
  - Usage model (space shared, I/O structure etc.)
  - Scalability of the machine (balance)

Porting/Developing Systems Software

- BG/L default environment is by necessity a minimal model
- Many more tools could be brought to the BG/L platform via the community
- Some new tools need to be developed
  - BGC provides a forum for discussions, priority setting and collaboration
- Understanding how to make BG/L class systems more usable could be a major goal of the CS community involved in the consortium

Applications Development

- BG/L and potential followon's offer dramatic capability for some applications
  - How many types of applications can take advantage of BG?
  - For those that can take advantage how far can these applications be scaled?
  - Can we characterize the successful and unsuccessful applications in ways that are predictive?
  - Can we identify quantitative factors in BG/L design that are primarily responsible for application appropriateness?
- What are the key systems software elements needed to advance relevant applications?

mcs

Preliminary BG Evaluations

Code	BlueGene Application	Strengths	Evaluator
LINPACK	Linear equation solving benchmark	70% of peak on 1024 nodes	IBM
UTMK	Unstructured mesh radiation transport	Strong scaling, some load balancing issues	LLNL
SAGE	Adaptive grid Eulerian hydrodynamics	Better scaling than large ASCI clusters	IBM
GP	Ab initio molecular dynamics	Extensive scaling analysis Ğno obstacles	LLNL
DD3d	3D dislocation dynamics	100x space and time scales possible	LLNL
sPPM	3D compressible hydrodynamics	Will scale easily to 64K nodes	LLNL
3D-FFT	Distributed Fast Fourier Transform	1024^3 est. 66% efficiency on 64K nodes	IBM
mpiBLAST	Computational gene discovery		LLNL
QMC	Quantum Monte Carlo materials	Global reductions do not dominate	LLNL
Nek5000	Fluid dynamics with turbulence	Highly optimized; proven scalability	ANL
pNEO	Neocortical seizure simulation	Potential for 1 B neurons on BG	ANL
AGCM	2D spectral atmospheric climate	Est. 2x faster than Earth Simulator on 64K	IBM

mcs

Port and Develop Important Systems Software to the BG/L Architecture

- Emerging and alternative programming models
  - UPC, CAF, CHARM, Java, etc.
- Numerical libraries
  - PETsc, Scalapack, R, Vis libraries, etc.
- Parallel Filesystems
  - PVFS, Lustre, etc.
- Parallel I/O and Grid software
  - GridFTP, Globus etc.
- Alternative kernels and services infrastructure
  - Linux on the compute nodes?
- Enabling the machine as a OS research testbed
  - Open source alternatives

mcs

Provide feedback to IBM on future systems related to BG/L

- Detailed feedback on what works and what doesn't
- Functional requests based on extensive usage on BG/L
- Performance of the networks
- Performance of the memory/caches
- Floating point performance
- Novel use of the second CPU
- Software architecture feedback
- Usage model feedback
- I/O architecture freedback
- Etc.

mcs

Organizational Structure for Providers and Users of the BG/L Systems

- We envision a plan that looks something like:
  - 6-8 Hardware partners with machines ranging from 1 to 10 racks
    - Consortium provides community model for support and common configs
  - 6-8 Systems software collaborators covering major tools and libraries
    - Consortium provides venue for interaction and collaboration
  - 10-20 applications collaborators spanning existing applications domains and major community codes
    - Consortium provides environment with experienced developers
  - 6-10 CS departments featuring BG related activities
    - Consortium provides outreach and training
  - 4-6 dedicated BG staff at multiple sites providing the glue and the engine for making this work
    - Consortium provides front line support and community interaction
  - BG/L user community of about 100-200 people
    - Consortium provides opportunity for leading edge science

mcs

Support the engagement of CS departments on BG related software projects

- We would like to investigate the feasibility of a workstation class machine that leverages BG/L technology that would provide a 32-64 node development environment front-ended by a commodity Linux workstation
- Targeting
  - Student projects for novel systems software
  - Develop the idea of the BG building block
  - Applications and tools porting environment
  - Desktop to Petaflops

mcs

Goals for the Kickoff Meeting

- Identify candidates for hardware purchases and determine if there is a critical mass interested in cooperatively purchasing hardware
  - Determine the degree of cooperation that is feasible
  - Determine the types of agency support that might be feasible
- Identify groups interested in porting applications and systems software and to prioritize those efforts
  - Figure out ways to accelerate this activity
- Organize sites interested in providing access to the community to BG/L systems
  - Develop a user and provider community
  - Organize outreach activities
- Determine interest level for BG/L follow on systems and how community involvement in system development might work
  - Is it feasible to get the community involved in the next generation design?

Breakout Questions

- Consortium Management
- Hardware Acquisition and Operation
- Systems Software Development and Porting
- Applications Porting and Development
- Fast Forward.. BG influences on petaflops paths

Consortium Management

1. Determine a management structure for the consortium.. EC, working groups..
2. Determine an outreach strategy
3. Funding opportunities for joint projects involving consortium members.
4. Interactions with agencies in support of evaluation and deployment of BG systems.
5. Issues regarding BG roadmap to petaflops..
6. Coordination of US BG activities with international efforts related to BG..
7. Series of workshops in sw, apps, porting etc.
8. Engaging CS departments..
9. Personal version of BG for development.

Hardware Acquisition and Operations

1. Strategy for leveraging partners for HW config and install.
2. Email, web, etc for supporting installations
3. Plan for coordinating bugs with llnl
4. Training.
5. Self help support model and interfaces to IBM
6. Standard configuration issues..
7. Sharing of experiences
8. Proving access to consortium members..
9. Supporting develop and testbed activities..

mcs

Systems Software

1. List of high priority SS items for porting..
2. Needs for accelerating SS for BGL etc.
3. Coordination mechanisms..
4. Stuff needed from IBM to facilitate SS porting and development
5. Strategies for encouraging more participation in SS dev.
6. Prospect of an all open source SW stack?
7. Requirements for additional file systems, scheduling options
8. Putting systems on the Grid? External IO.
9. SS Infrastructure for using the systems as a testbed..

Applications

1. List of apps in good shape for porting to BGL and who is doing them
2. Apps of interest to consortium members..Who is interested/committed to which..
3. Apps we think should be targeted in future meetings as possibly good targets for BG
4. Requirements for porting and performance studies..
5. Scalability shots.. Planning for limited access..
6. Prioritized list of SS needed.. Libraries.. OS function shipping etc.
7. Support needed from other consortium teams..
8. List of meetings and workshops needed.
9. Tools needed to make this work..

Fast Forward.. Paths to Petaflops

1. What types of feedback from the consortium would be the most helpful for BG petaflops?
2. Access to simulators and other tools that will enable the consortium to participate in evaluating. BG petaflops design points..
3. Critical tests and experiments/measurements that BG consortium members could help with.
4. Rough timetable for feedback of various types for influencing the design..
5. Systems software issues that might be important for BG petaflops timeframe..

mcs